ACADEMIC ACHIEVEMENT PREDICTION MODEL
USING NEURAL NETWORKS

A thesis submitted to the Graduate School in partial
fulfillment of the requirements for the degree
Master of Science (Information Technology),
Universiti Utara Malaysia

by
Normaziah Binti Abdul Rahman

(c) Normaziah binti Abdul Rahman, 2002. All rights reserved
PERAKUAN KERJA KERTAS PROJEK
(Certification of Project Paper)

Saya, yang bertandatangan, memperakukan bahawa
(I, the undersigned, certify that)

NORMAZIAH BINTI ABDUL RAHMAN

calon untuk Ijazah
(candidate for the degree of) Master of Science (Information Technology)
telah mengemukakan kertas projek yang bertajuk
(has presented his/her project paper of the following title)

ACADEMIC ACHIEVEMENT PREDICTION MODEL USING

NEURAL NETWORKS

seperti yang tercatat di muka surat tajuk dan kulit kertas projek
(as it appears on the title page and front cover of project paper)
bahawa kertas projek tersebut boleh diterima dari segi bentuk serta kandungan,
dan meliputi bidang ilmu dengan memuaskan.
(that the project paper acceptable in form and content, and that a satisfactory
knowledge of the field is covered by the project paper).

Nama Penyelia
(Name of Supervisor) : Prof. Madya Fadzilah Siraj

Tandatangan
(Signature) : [Signature]

Tarikh
(Date) : 23/05/2002

FADZILAH SIRAJ
Associate Professor
School of Information Technology
Universiti Utara Malaysia
PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirement for a post graduate degree from Universiti Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by my supervisor or, in their absence, by the Dean of the Graduate School. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole part, should be addressed to:

Dean of Graduate School
Universiti Utara Malaysia
06010 UUM Sintok
Kedah Darul Aman
ABSTRACT (BAHASA MELAYU)


Penyelesaian berasaskan rangkaian neural dengan menggunakan Multi Layer Perceptron (MLP) dan algoritma rambatan balik telah digunakan di dalam pembangunan model ini. Sejumlah 248 sampel data telah dikutip daripada pelajar Diploma Teknologi Maklumat dan Multimedia dan sampel ini telah dilatih dan diuji menggunakan model ini. Ketepatan ramalan bagi latihan sebanyak 90% dan ujian sebanyak 83.33% telah dicapai menggunakan model ini.

Analisa data menunjukkan terdapat hubung-kait antara pembolehubah input, yang mana ia terdiri daripada umur, jantina, lokasi sekolah, aliran pelajaran dan keputusan untuk matapelajaran tertentu iaitu English, matematik, Sains, Fizik dan Matematik Tambahan, dengan pembolehubah output. Keputusan juga menunjukkan rangkaian neural mempunyai potensi di dalam bidang pendidikan.
ABSTRACT (ENGLISH)

This study aims to develop the academic achievement prediction (ACP) model using Neural Networks. It is capable of predicting the student's result in Programming I (C Language) subject for Kolej Agama Sultan Zainal Abidin, Kuala Terengganu. This model allows the system administrator to train and normalize data as well as trains. Once the model has been established by the administrator, the future student achievement can be forecast by the model. The system can predict the result of Programming I subject based on the student's background during the Sijil Pelajaran Malaysia (SPM) examination.

A neural network technique, using Multi Layer Perceptron (MLP) and back propagation algorithm was employed. A total of 248 data samples from Information Technology and Multimedia Diploma students were collected, trained and tested using this model. A training prediction of 90% accuracy and testing prediction of 83.33% accuracy were achieved using this model.

The analysis of the data shows a reasonably strong correlation between the input variables, which consist of age, gender, school location, subject stream and result for a certain subjects: English, Mathematics, Science, Physics and Additional Mathematics, with the targeted output variable. The results also indicate that neural network has a potential to be used for education planning.
ACKNOWLEDGEMENTS

Throughout this project, I was fortunate to have had the help and contributions of my supervisors, Prof. Madya Fadzilah Siraj. I would like to extend my thanks to my beloved husband, Fazli and my daughter, Aisyah for the courage and understanding. This project would not have been possible without their encouragement, support and guidance.
TABLE OF CONTENTS

PERMISSION TO USE i
ABSTRACT (BAHASA MELAYU) ii
ABSTRACT (ENGLISH) iii
ACKNOWLEDGMENTS iv
LIST OF TABLES viii
LIST OF FIGURES x

CHAPTER ONE : INTRODUCTION

1.1 Introduction 1
1.2 Problem Definition 3
1.3 Purposes and Scopes Of The Study 4
1.4 Research Questions 4

CHAPTER TWO : LITERATURE REVIEW

2.1 Academic Achievement 6
2.2 Prediction Model And Forecasting Model 8
2.3 Artificial Intelligent Approach 10

CHAPTER THREE : ARTIFICIAL NEURAL NETWORK

3.1 The Concept of Neural Network 16
3.2 Neural networks versus conventional computers 19
3.3 The Analogy to the Brain 20
  3.3.1 The Biological Neuron 21
  3.3.2 The Artificial Neuron 22
3.4 Architecture of Neural Networks
  3.4.1 Feed-forward networks 23
  3.4.2 Feedback networks 24
  3.4.3 Network layers 24
3.5 The Learning Process 25
3.6 Multilayer Perceptrons 28

CHAPTER FOUR : METHODOLOGY

4.1 Identifying Attributes 36
4.2 Data Collection Procedures 37
4.3 Pre-Processing Techniques 38
4.4 System Development 40
  4.4.1 Database Design 40
  4.4.2 Context Diagram 41
  4.4.3 Data Flow Diagram 42
  4.4.4 Menu Structure 44
  4.4.5 Flow Chart 44
4.5 Training Pseudo Code 49
  4.5.1 Initialize Weight 49
  4.5.2 Feed Forward 50
  4.5.3 Backpropagation of error 51

CHAPTER FIVE : RESULT

5.1 Training And Testing Results 54
5.2 Identifying The Most Suitable Hidden Unit (HU) 55
5.3 Identifying The Most Suitable Value Of Learning Rate 58
5.4 Identifying The Most Suitable Of Momentum Rate 61
5.5 Identifying The Most Suitable Of Stopping Criteria 63
5.6 Identifying The Most Suitable Number Of Epoch 66
5.7 Neural Network Model 68

CHAPTER SIX: CONCLUSION AND SUGGESTION

6.1 Conclusions 69
6.2 Suggestions 70

REFERENCES

APPENDIXES

APPENDIX A
APPENDIX B
APPENDIX C
APPENDIX D
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table No.</th>
<th>Name of Table</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>The list of attributes</td>
<td>36</td>
</tr>
<tr>
<td>4.2</td>
<td>A sample data summary</td>
<td>37</td>
</tr>
<tr>
<td>4.3</td>
<td>Normalized technique</td>
<td>38</td>
</tr>
<tr>
<td>4.4</td>
<td>Target values</td>
<td>40</td>
</tr>
<tr>
<td>4.5</td>
<td>Normalized data</td>
<td>40</td>
</tr>
<tr>
<td>4.6</td>
<td>Database design</td>
<td>41</td>
</tr>
<tr>
<td>5.1</td>
<td>Starting values of parameter before training process</td>
<td>55</td>
</tr>
<tr>
<td>5.2</td>
<td>Percentage accuracy of different numbers of Hidden Units.</td>
<td>56</td>
</tr>
<tr>
<td>5.3</td>
<td>The results of using HU = 6 and HU = 8 using various weight seeds</td>
<td>57</td>
</tr>
<tr>
<td>5.4</td>
<td>Percentage accuracy of different values of learning rate</td>
<td>59</td>
</tr>
<tr>
<td>5.5</td>
<td>The results of using learning rate = 0.8 and learning rate = 1.0 using various weight seeds</td>
<td>60</td>
</tr>
<tr>
<td>5.6</td>
<td>Percentage accuracy of different values of momentum rate</td>
<td>61</td>
</tr>
<tr>
<td>5.7</td>
<td>The results of using momentum = 0.3 and momentum = 0.8 using various weight seeds</td>
<td>62</td>
</tr>
<tr>
<td>5.8</td>
<td>Percentage accuracy of different percentage of stopping criteria</td>
<td>64</td>
</tr>
<tr>
<td>5.9</td>
<td>The results of using stopping criteria = 80% and stopping criteria = 85% using various weight seeds</td>
<td>65</td>
</tr>
<tr>
<td>5.10</td>
<td>Percentage accuracy of different number of epoch.</td>
<td>67</td>
</tr>
</tbody>
</table>
5.11 Final results for each parameter 68
6.1 Comparison of result between logistic regression model and multi
layer perceptron.
<table>
<thead>
<tr>
<th>Figure No.</th>
<th>Figure Name</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>A simplified biological neuron</td>
<td>21</td>
</tr>
<tr>
<td>3.2</td>
<td>The basics of an artificial neuron</td>
<td>22</td>
</tr>
<tr>
<td>3.3</td>
<td>An example of a simple feedforward network</td>
<td>23</td>
</tr>
<tr>
<td>3.4</td>
<td>Architecture of multi-layer perceptron</td>
<td>29</td>
</tr>
<tr>
<td>3.5</td>
<td>The sigmoid curve</td>
<td>31</td>
</tr>
<tr>
<td>3.6</td>
<td>Training procedure for multi-layer perceptron network</td>
<td>34</td>
</tr>
<tr>
<td>4.1</td>
<td>Context diagram</td>
<td>42</td>
</tr>
<tr>
<td>4.2</td>
<td>Data flow diagram</td>
<td>43</td>
</tr>
<tr>
<td>4.3</td>
<td>Menu Structure</td>
<td>44</td>
</tr>
<tr>
<td>4.4</td>
<td>Main flow chart</td>
<td>45</td>
</tr>
<tr>
<td>4.5</td>
<td>MLP flow chart</td>
<td>46</td>
</tr>
<tr>
<td>4.6</td>
<td>Student registration flow chart</td>
<td>47</td>
</tr>
<tr>
<td>4.7</td>
<td>The prediction flow chart</td>
<td>48</td>
</tr>
<tr>
<td>5.1</td>
<td>Percentage accuracy for the various numbers of Hidden Units</td>
<td>56</td>
</tr>
<tr>
<td>5.2</td>
<td>Percentage accuracy VS seed number for different numbers of Hidden Unit</td>
<td>58</td>
</tr>
<tr>
<td>5.3</td>
<td>Percentage accuracy of various learning rate for training and testing data</td>
<td>59</td>
</tr>
<tr>
<td>5.4</td>
<td>Percentage accuracy VS seed number for different values of learning rate</td>
<td>60</td>
</tr>
<tr>
<td>5.5</td>
<td>Percentage accuracy of various momentum rate for training and testing data</td>
<td>62</td>
</tr>
<tr>
<td>5.6</td>
<td>Percentage accuracy VS seed number for different values of momentum rate</td>
<td>63</td>
</tr>
<tr>
<td>5.7</td>
<td>Percentage accuracy of various stopping criteria for training and testing data</td>
<td>64</td>
</tr>
<tr>
<td>5.8</td>
<td>Percentage accuracy VS seed number for different values of stopping criteria</td>
<td>66</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>5.9</td>
<td>Percentage accuracy of various number of epoch for training and testing data</td>
<td>67</td>
</tr>
<tr>
<td>5.10</td>
<td>Multi layer perceptron model</td>
<td>69</td>
</tr>
</tbody>
</table>
CHAPTER ONE

INTRODUCTION

1.1 Introduction

As technology continuously progresses, methodologies evolve to enhance our abilities to perform arduous tasks more expediently. Utilizing modern computing technologies not only makes completing tasks more efficient, but also often achieves a higher degree of accuracy than do humans. Neural network has emerged as a very popular area of research, both from the design and the usage points of view. There is considerable research emphasis on designing better and more efficient neural networks, more powerful "learning algorithms", better transfer functions. On the other hand, there is a great amount of academic interest in the applications of neural networks. In addition, there is a significant volume of research on neural networks in the engineering and science literature. There also exists a reasonable body of neural network research as related to business.
The contents of the thesis is for internal user only
REFERENCES:

A Cripps. (1996). “Predicting grade point average using clustering and ANNs”.
Proceedings of World Congress on Neural Networks, 486-490.


predictors of school achievement ". Journal of Educational Psychology, 63, (4),
398-404

King College of London.

URL: http://ksi.cpsc.ucalgary.ca/articles/BRETAM/InfSci/

C. Van Nelson dan L.W. Henriksen. (1994). "Using neural net technology to enhance the
efficiency of a computer adaptive testing application ". Mid – Western

Coleman, J. S., E. Q. Campbell, C. J. Hobson, J. McPartland, A. M. Mood, F. D.

state policy evidence ". Education Policy Analysis Archives 8 (1).


Steve Beveridge. (2001)." Forecasting for Planners and Managers "
URL: http://courses.bus.ualberta.ca/mgtsc405-beveridge

URL : http://www.personal.psu.edu/staff/s/m/smc258/KB/CRBarticle.htm


